synthesized only in the presence of both rtTA and tetracycline derivatives. Accordingly, offspring containing both transgenes develop tumors in the cells expressing the gene with the retroviral insertion upon administration of tetracycline derivatives.

This approach affords a number of advantages over previous methodologies for the generation of transgenic mouse models for neoplasia (Quaife et al., Cell **48**:1023-1034 (1987); Sinn et al., Cell **49**:465-475 (1987); Jat et al., Proc. Natl. Acad. Sci. USA **88**:5096-5100 (1991); and Sandmoller et al., Cell Growth and Diff. **6**:97-103 (1995)). First, the method is not limited by the restricted spectrum of available tissue-specific promoters. And, second, the oncogenic state is not constitutive, but is conditional; the neoplastic transformation of a normal mouse tissue is initiated only in the presence of tetracycline derivatives, making the system more amenable to analysis. Animals generated by this method provide information about the types of oncogenes which play roles in particular cell types, and may also be used as animal models to screen anti-cancer therapies.

MAGEKO allows establishment of conditional immortal cell lines of any desired type. Once available, animals with developed tumors of desired cellular origin (produced as described above) are an immediate source of tumor cell lines. In the alternative, immortal cell lines can be established from these animals prior to tumor development, simply by isolating the desired cells from the animals and culturing in vitro in the presence of tetracycline derivatives.

Such cell lines provide a valuable reagent for high throughput drug screening procedures to identify compounds which affect the gene with the retroviral insertion. In particular, since these cell lines express the target gene and also constitutively synthesize GFP, the cell lines are, for example, GFP^+, β-GAL^+ (if β-GAL is the reporter gene). Any drug that specifically affects the gene in question produces GFP^+, β-GAL^- cells.

(III) LOK

The retroviral vectors described above are used to construct libraries of ES cells containing knock outs in endogenous genes or "LOKs". To produce these libraries, the vectors are introduced by infection into ES cells to obtain insertions, on the average, every 1 Kb in the genome. The LOK preferably consists of thirty million such insertions, each carrying an independent provirus. The complexity of an LOK is high enough that most mouse genes should statistically be hit at least once by an independent retroviral integration event.

Following infection of ES cells with the retroviral vectors, transduced cells expressing the visual marker (for example, GFP) are selected by FACS analysis, and the cells are distributed in multi-well plates. The contents of combinations of wells are then pooled, subsequent to duplicate formation and storage of the replica, and appropriate matrices are generated to facilitate assignment of a specific cell to a particular well. Several proposed and established pooling strategies are available for the generation of the desired matrix to screen the LOK (see, for example, Zwaal et al., Proc. Natl. Acad. Sci. USA 90:7431-7435 (1993); Evans and Lewis, Proc. Natl. Acad. Sci. USA **86**:5030-5034 (1989); Green and Olson, Proc. Natl. Acad. Sci. USA **87**:1213-1217 (1990); Kwiatkowski et al., Nucleic Acids Res. **18**:7191-7192 (1990); and Barillot et al., Nucleic Acids Res. **19**:6241-6247 (1991)).

(IV) KIS

The present invention also includes a gene Knock out Identification System (or "KIS"). According to this aspect of the invention, genomic DNA, including the integrated nucleic acids of the retroviral vectors, are isolated from the pooled ES cells of the LOK and are fragmented. These fragments are then circularized and amplified by inverted PCR (see, for example, Ochman et al., Genetics **120**:621-625 (1988); and Triglia et al., Nucleic Acids Res. **16**:8186 (1988)), using primers which hybridize to the retroviral vector sequences but which are not present in the mouse genome; this method

has been successfully applied to the detection of retroviral insertions in the *Zebrafish* genome (Allende, Genes Dev. **10**:3141-3155 (1996)), *P* element insertions in the *Drosophila* genome (Dalby et al., Genetics **139**:757-766 (1995)) and transposon insertions in the *Arabidopsis* genome (Sundaresan et al., Genes Dev. **9**:1797-1810 (1995)). Alternatively, modifications of inverse PCR, such as oligo-cassette mediated PCR (see, for example, Rosenthal and Jones, Nucleic Acids Res. **18**:3095-3096 (1990)) or ligation mediated PCR (see, for example, Mueller and Wold, Science **246**:780-786 (1989)), may be used.

Genomic DNA fragments, once amplified, are transferred to hybridization supports, generating an ordered array of genomic DNA flanking the provirus. Labelled DNA from a gene of interest is then hybridized to the pooled genomic DNA, and a positive signal leads to the rapid identification of the desired ES cell clone.

Alternatively, detection of a retroviral integration site may be accomplished by direct sequencing of the amplified DNA of an ES clone; this approach, however, requires the isolation of single clones of ES cells and is preferably used only for a subset of the generated clones. In another alternative approach, an integration site may be determined by sequence detection using a positional oligonucleotide probing technique (POP), a method which is ideal for the processing of limited sequence information in parallel. According to this technique, all possible oligonucleotides of a specific length are synthesized in a high density array (such as an Affymetrix chip (see, for example, Lipshutz et al., BioTechniques **19**:442-447 (1995)) and hybridized to the amplified DNA from ES cells. The POP technique is based on generating sequence information for an unknown region of nucleic acid (i.e., the genomic DNA), which is linked to a known sequence (i.e., a portion of the retroviral vector). Because retroviral integration is precise and results in the integration of a viral LTR within the genomic DNA, the LTR sequence is a preferred sequence for designing oligonucleotide probes. For example, oligonucleotides that contain 8 bases corresponding to the tip of the LTR and nine